AD-A117 263 TEXAS A AND M UNIV COLLEGE STATION INST OF STATISTICS F/G 12/1
MAXIMUM ENTROPY INTERPRETATION OF AUTOREGRESSIVE SPECTRAL DENSI--ETC(U)
JUN 82 E PARZEN DAAG29-80-C-0070
UNCLASSIFIED TR-N-30 NL

1 OF 1
AD A117263

END
DATE
FILMED
8 82
DTIC

(12)

TEXAS A&M UNIVERSITY

COLLEGE STATION, TEXAS 77843-3143

INSTITUTE OF STATISTICS
713 - 845-3141

AD A117263

MAXIMUM ENTROPY INTERPRETATION OF AUTOREGRESSIVE SPECTRAL DENSITIES

by Emanuel Parzen

Institute of Statistics

Texas A&M University

DTIC
JUL 21 1982
H

Technical Report No. N-30

June 1982

Texas A&M Research Foundation
Project No. 4226T

"Multiple Time Series Modeling and
Time Series Theoretic Statistical Methods"

Sponsored by the Office of Naval Research

Professor Emanuel Parzen, Principal Investigator

DTIC FILE COPY

Approved for public release; distribution unlimited.

82 ... 011

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER N-30	2. GOVT ACCESSION NO. AD-A117263	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) MAXIMUM ENTROPY INTERPRETATION OF AUTOREGRESSIVE SPECTRAL DENSITIES		5. TYPE OF REPORT & PERIOD COVERED Technical
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Emanuel Parzen		8. CONTRACT OR GRANT NUMBER(s) ONR N00014-82-MP-2001 ARU DAAG 29-80-C-0070
9. PERFORMING ORGANIZATION NAME AND ADDRESS Texas A&M University Institute of Statistics College Station, TX 77843		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS		12. REPORT DATE June 1982
		13. NUMBER OF PAGES
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

NA

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Maximum entropy spectral estimators, autoregressive spectral density, exponential probability density, entropy, cross-entropy, information divergence.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A new proof is given of the maximum entropy characterization of autoregressive spectral densities as models for the spectral density of a stationary time series. The new proof is presented in parallel with a proof of the maximum entropy characterization of exponential models for probability densities. Concepts of entropy, cross-entropy, and information divergence are defined for probability densities and for spectral densities.

DD FORM 1473, 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-LF-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

DTIC COPY INSPECTED 2

Accession For
NTIS GRA&I
DTIC TAB
Unannounced
Justification
By
Distribution/
Availability Codes
Dist | Avail and/or Special
A

The use of autoregressive spectral densities as exact models, and as approximating models, [see Akaike (1974, Parzen (1974), Priestley (1981)] for true spectral densities is often questioned by sceptical statisticians on the ground that their use in general is ad hoc and without theoretical justification. A possible answer to this criticism is provided by entropy concepts. This paper presents a new proof of the maximum entropy character of autoregressive spectral densities, first proved by VanDenBos (1971) following work of Burg (1968).

1. Time series background

The definitions and notation we adopt for the functions used to describe a zero mean stationary Gaussian discrete parameter time series $Y(t)$, $t=0, \pm 1, \ldots$ are as follows.

A "time domain" specification of the probability law of $Y(\cdot)$ is provided by the covariance function.

$$R(v) = E[Y(t)Y(t+v)], \quad v=0, \pm 1, \pm 2, \ldots;$$

or by the variance $R(0)$ and the correlation function

$$\rho(v) = \frac{R(v)}{R(0)} = \text{Corr}\,[Y(t), Y(t+v)].$$

To define spectral (frequency) domain specification of the probability law of $Y(\cdot)$ we first assume summability of $R(\cdot)$ and $\rho(\cdot)$. The Fourier transforms of $R(v)$ and $\rho(v)$ are called

the <u>power spectrum</u> $S(\omega)$ and <u>spectral density function</u> $f(\omega)$ respectively, and are defined by

$$S(\omega) = \sum_{v=-\infty}^{\infty} e^{-2\pi i v\omega} R(v), \quad 0\leq\omega\leq 1;$$

$$f(\omega) = \sum_{v=-\infty}^{\infty} e^{-2\pi i v\omega} \rho(v), \quad 0\leq\omega\leq 1.$$

A spectral density is called an autoregressive spectral density when it can be expressed in the form of eq. (3.9) below. They are used for nonparametric estimation of spectral densities, and for time series model identification.

Parzen (1982) proposes that it is useful in practice to distinguish qualitatively between three types of time series:

no memory: white noise,

short memory: stationary and ergodic,

long memory: non-stationary or non-ergodic.

A <u>no-memory</u> or <u>white noise</u> time series is a stationary Gaussian time series satisfying either of the equivalent conditions: $\rho(v) = 0$ for $v>0$; $f(\omega) = 1, 0\leq\omega\leq 1$.

A <u>short memory</u> time series is a stationary time series possessing a summable correlation function $\rho(v)$ and a spectral density $f(\omega)$ which is bounded above and below in the sense that the dynamic range of $f(\omega)$

$$DR(f) = \left\{\max_{0\leq\omega\leq 1} f(\omega)\right\} \div \left\{\min_{0\leq\omega\leq 1} f(\omega)\right\}$$

satisfies $1<DR(f)<\infty$. Then $f(\omega)$ can be shown to be representable as the limit of a sequence of autoregressive spectral densities $f_m(\omega)$.

A long memory time series is one which is neither no memory nor short memory; alternativey, a long memory time series is one which is non-stationary or non-ergodic. It usually has components representing cycles or trends.

2. Entropy and exponential models

The notion of entropy in statistics is usually first defined for a discrete distribution with probability mass function $p(x)$. The entropy of this distribution, denoted $H(p)$, is defined by

$$(1) \qquad H(p) = -\sum_{x} p(x) \log p(x)$$

For the distribution of a continuous real valued random variable X, with probability density function $f(x)$, entropy is defined (analogously or formally) by

$$(2) \qquad H(f) = -\int_{-\infty}^{\infty} f(x) \log f(x)\, dx$$

A concept closely related to entropy is information divergence $I(f;g)$ between two probability densities $f(x)$ and $g(x)$, defined by

$$(3) \qquad I(f;g) = \int_{-\infty}^{\infty} \{-\log \frac{g(x)}{f(x)}\} f(x)\, dx$$

The measure (3) is called by statisticians the Kullback-Liebler number because it was introduced into statistics in Kullback and Liebler (1951). It seems that a more correct name for (3) would be the Kullback number, as the concept of the use of these numbers in statistical inference, as in Kullback (1959), is entirely due to Kullback.

One should note that $I(f;g)$ equals minus the generalized entropy $H(f|g)$ defined by

(4) $$H(f|g) = \int_{-\infty}^{\infty}\{-\frac{f(x)}{g(x)} \log \frac{f(x)}{g(x)}\} g(x)\, dx$$

Another fundamental concept is cross-entropy defined by

(5) $$H(f;g) = \int_{-\infty}^{\infty} \{-\log g(x)\} f(x)\, dx.$$

Note that $H(f) = H(f;f)$.

Information divergence is expressed in terms of cross-entropy and entropy by

(6) $$I(f;g) = H(f;g) - H(f)$$

Important Information Inequality:

(7) $$I(f;g) \geq 0$$

with equality if and only if $f = g$; consequently

(8) $$H(f) \leq H(f;g)$$

Some applications of entropy in probability and statistical modeling are now described.

The method of maximum likelihood parameter estimation can be described abstractly as follows. One introduces a parametric family of probability densities $f_\theta(x)$, indexed by a vector

parameter $\theta = (\theta_1, \ldots, \theta_k)$. Suppose there is a true parameter value $\bar{\theta}$ in the sense that the true probability density $f(x) = f_{\bar{\theta}}(x)$. Then $\bar{\theta}$ satisfies

$$(10) \qquad H(f) = H(f;f_{\bar{\theta}}) = \min_{\theta} H(f;f_{\theta}).$$

To estimate $\bar{\theta}$ from data, one forms an estimator $\tilde{H}(f;f_{\theta})$ of $H(f;f_{\theta})$ and defines an estimator $\hat{\theta}$ of $\bar{\theta}$ by

$$(11) \qquad \tilde{H}(f;f_{\hat{\theta}}) = \min_{\theta} \tilde{H}(f;f_{\theta}).$$

The estimator $\tilde{H}(f;f_{\theta})$ could be of the form

$$(12) \qquad \tilde{H}(f;f_{\theta}) = H(\tilde{f};f_{\theta})$$

for a suitable raw estimator $\tilde{f}(x)$ of $f(x)$.

The parametric families of probability densities $f_{\theta}(x)$ are often derived axiomatically using a maximum entropy principle. Natural Exponential models: A parametric family of probability densities $f_{\theta}(x)$ is said to obey a natural exponential model when it is of the form

$$(13) \qquad \log f_{\theta}(x) = \sum_{j=1}^{k} \theta_j T_j(x) - \Psi(\theta_1, \ldots, \theta_k)$$

where

$$(14) \qquad \Psi(\theta_1, \ldots, \theta_k) = \log \int_{-\infty}^{\infty} dx \exp \sum_{j=1}^{k} \theta_j T_j(x) \quad ,$$

Natural exponential models are maximum entropy probability densities in the sense of the following theorem [see Guiasu (1977) and Kagan, Linnik, and Rao (1973), p. 409]. Fix k

functions $T_j(x)$, $j=1,2,\dots,k$, and k real numbers $\tau_1, \tau_2, \dots, \tau_k$ such that there exists probability densities f(x) satisfying

$$(15) \qquad \int_{-\infty}^{\infty} T_j(x)\, f(x)\, dx = \tau_j, \quad j=1,\dots,k.$$

The density with maximum entropy H(f) among these densities is of the form (13) where $\theta_1,\dots,\theta_k$ are chosen to satisfy

$$(16) \qquad \int_{-\infty}^{\infty} T_j(x)\, f_\theta(x)\, dx = \tau_j\,, \quad j=1,\dots,k.$$

The aim of this paper is to present a new proof of the maximum entropy character of autoregressive spectral densities which is analogous to the simple proof of the maximum entropy character of exponential models for probability densities. We recall the latter. Verify that for any f(x) satisfying the moment constraints (15)

$$(17) \qquad H(f;f_\theta) = \Psi(\theta_1,\dots,\theta_k) - \sum_{j=1}^{k} \theta_j \tau_j = H(f_\theta),$$

and therefore

$$(18) \qquad H(f) \le H(f;f_\theta) = H(f_\theta).$$

Thus the maximum entropy is achieved by $f_\theta(x)$.

3. Entropy of spectral density functions

To extend entropy concepts to short memory stationary zero mean Gaussian time series, define the information divergence for a sample $Y(t)$, $t=1,2,\ldots,T$ as a function of the true probability density f of the sample, and a model g for f. We define

$$(1) \qquad I(f;g) = \lim_{T\to\infty} I_T(f;g) \quad ,$$

$$(2) \qquad I_T(f;g) = \frac{-1}{T} E_f \left[\log \frac{g(Y(1),\ldots,Y(T))}{f(Y(1),\ldots,Y(T))} \right]$$

It should be noted that we are using the notation f and g with a variety of meanings. For a Gaussian zero mean stationary time series, the probability density of the sample is specified by the spectral densities $f(\omega)$ of the true distribution and $g(\omega)$ of the model. The arguments of the information divergence $I(f;g)$ indicate spectral densities in the following discussion. Pinsker (1963) derives the following very important formula:

$$(3) \qquad I(f;g) = \frac{1}{2} \int_0^1 \left\{\frac{f(\omega)}{g(\omega)} - \log \frac{f(\omega)}{g(\omega)} - 1\right\} d\omega$$

Since $u - \log u - 1 \geq 0$ for all u, I has two of the properties of a distance: $I(f;g) \geq 0$, $I(f;f) = 0$. However I does not satisfy the triangle-inequality.

We define the cross-entropy of spectral density functions $f(\omega)$ and $g(\omega)$ by

$$(4) \qquad H(f;g) = \frac{1}{2}\int_0^1 \{\log g(\omega) + \frac{f(\omega)}{g(\omega)}\}\, d\omega$$

The entropy of f is

$$(5) \qquad H(f) = H(f;f) = \frac{1}{2}\int_0^1 \{\log f(\omega) + 1\}\, d\omega$$

Information divergence can be expressed

$$(6) \qquad I(f;g) = H(f;g) - H(f)$$

Hence

$$(7) \qquad H(f) \leq H(f;g)$$

An approximating autoregressive spectral density of order m, denoted $\bar{f}_m(\omega)$, to a spectral density $f(\omega)$ is defined by

$$(8) \qquad H(f;\bar{f}_m) = \min_{f_m} H(f;f_m)$$

where the minimization is over all f_m of the form

$$(9) \qquad f_m(\omega) = \sigma_m^2 \, |g_m(e^{2\pi i\omega})|^{-2} ,$$

$$(10) \qquad g_m(z) = 1 + \alpha_m(1)\, z + \ldots + \alpha_m(m)\, z^m$$

One may verify that

$$(11) \qquad H(f;f_m) = \frac{1}{2} \{\log \sigma_m^2 + \frac{1}{\sigma_m^2} \int_0^1 g_m|(e^{2\pi i\omega})|^2 f(\omega)d\omega\}$$

The coefficients $\bar{\sigma}_m^2$, $\bar{\alpha}_m(1),\ldots,\bar{\alpha}_m(m)$ of the minimum cross-entropy approximating autoregressive spectral density satisfy

$$(12) \qquad \bar{\sigma}_m^2 = \int_0^1 |\bar{g}_m(e^{2\pi i\omega})|^2 f(\omega)\, d\omega$$

$$= \int_0^1 \bar{g}_m(e^{2\pi i\omega})\, f(\omega)\, d\omega$$

$$= \min_{g_m} \int_0^1 |g_m\, (e^{2\pi i\omega})|^2 f(\omega)\, d\omega,$$

$$(13) \qquad \int_0^1 \bar{g}_m(e^{2\pi i\omega}) e^{-2\pi ik\omega} f(\omega)\, d\omega \quad |$$

$$= \sum_{j=0}^{m} \bar{\alpha}_m(j)\, \rho(j-k) = 0, \quad k=1,2,\ldots,m$$

Further

$$(14) \qquad H(f;\bar{f}_m) = \frac{1}{2} \{\log \bar{\sigma}_m^2 + 1\} = H(\bar{f}_m)$$

The autoregressive spectral density $\bar{f}_m(\omega)$ can be derived axiomatically using a maximum entropy principle.

<u>Theorem</u>: The spectral density with maximum entropy among all spectral densities $f(\omega)$ satisfying the constraints

$$(15) \qquad \int_0^1 e^{2\pi i\omega j} f(\omega)\, d\omega = \rho(j), \quad j=1,2,\ldots,m$$

for m specified correlation coefficients $\rho(1),\ldots,\rho(m)$ is $\bar{f}_m(\omega)$ whose coefficients are determined by (12) and (13).

Proof: It may be verified that $\bar{f}_m(\omega)$ satisfies the constraints (15), and (14) holds for any $f(\omega)$ satisfying (15). Since

$$(16) \qquad H(f) \leq H(f;\bar{f}_m) = H(\bar{f}_m),$$

it follows that $\bar{f}_m$ has maximum entropy among all spectral densities satisfying the constraints (15).

A proof of this theorem based on prediction theory (due to Akaike), is given in Priestley (1981), p. 605. Our proof has the attraction of emphasizing the parallels between exponential densities and autoregressive densities.

4. <u>Extension to entropy of density-quantile functions</u>.

Parzen (1979) uses autoregressive densities to model quantile density functions

$$q(u) = \{F^{-1}(u)\}' = \{f(F^{-1}(u)\}^{-1}$$

The estimators derived may be shown to be maximum entropy estimators under the constraints imposed. This follows from

the fact that the entropy of a probability density function f(x) can be expressed

$$H(f) = \int_0^1 \log q(u)\, du = \int_0^1 -\log f(F^{-1}(u))\, du.$$

These integrals are defined to be the entropy of the quantile density function and density-quantile function respectively.

REFERENCES

Akaike, H. (1974) A new look at the statistical model identification, IEEE Trans. Autom. Contr., AC-19, 716-723.

Burg, J. P. (1968) A new analysis technique for time series data, NATO Advanced Study Institute on Signal Processing. Enschede, The Netherlands

Guiasu, S. (1977) Information Theory and Applications, McGraw Hill: New York.

Kagan, A.M., Linnik. Y. V., and Rao, C. R. (1973) Characterization Problems in Mathematical Statistics, Wiley: New York.

Kullback, S. (1959) Information Theory and Statistics, Wiley: New York.

Kullback, S. and Leibler, R. A. (1951) On information and sufficiency. Ann. Math. Statist. 22, 79-86.

Parzen, E. (1974) Some recent advances in time series modeling IEEE Trans. Automat. Contr., AC-19, 723-730.

_________. (1979) Nonparametric statistical data modeling. Journal of the American Statistical Association, 74, 105-131.

_________. (1982). Time Series Model Identification and Prediction Variance Horizon, Applied Time Series Analysis II, ed. D. Findley, Academic Press: New York, 415-447.

Pinsker, M. (1963) Information and Information Stability of Random Variables, Holden Day: San Francisco.

Priestley, M. B. (1981) Spectral Analysis and Time Series, Academic Press: London.

VanDenBos, A. (1971) Alternative interpretation of maximum entropy spectral analysis, IEEE Trans. Inform. Theory, IT-17, 493-494.